

7.7.6 Data extraction for time-to-event outcomes

Time-to-event outcomes are described in Chapter 9, Section 9.2.6, and their meta-analysis is discussed in Chapter 9, Section 9.4.9. Meta-analysis of time-to-event data commonly involves obtaining individual patient data from the original investigators, re-analysing the data to obtain estimates of the log hazard ratio and its standard error, and then performing a meta-analysis (see Chapter 18). Conducting a meta-analysis using summary information from published papers or trial reports is often problematic as the most appropriate summary statistics are typically not presented. Two approaches can be used to obtain estimates of log hazard ratios and their standard errors, for inclusion in a meta-analysis using the generic inverse variance methods, regardless of whether individual patient data or aggregate data are being used. For practical guidance, review authors should consult Tierney et al. (Tierney 2007).

In the first approach an estimate of the log hazard ratio can be obtained from statistics computed during a log-rank analysis. Collaboration with a knowledgeable statistician is advised if this approach is followed. The log hazard ratio (experimental relative to control) is estimated by $(O - E)/V$, which has standard error $1/\sqrt{V}$, where O is the observed number of events on the experimental intervention, E is the log-rank expected number of events on the experimental intervention, $O - E$ is the log rank statistic and V is the variance of the log-rank statistic. It is therefore necessary to obtain values of $O - E$ and V for each study.

These statistics are easily computed if individual patient data are available, and can sometimes be extracted from quoted statistics and survival curves (Parmar 1998, Williamson 2002). Alternatively, use can sometimes be made of aggregated data for each intervention group in each trial. For example, suppose that the data comprise the number of participants who have the event during the first year, second year, etc., and the number of participants who are event free and still being followed up at the end of each year. A log-rank analysis can be performed on these data, to provide the $O - E$ and V values, although careful thought needs to be given to the handling of censored times. Because of the coarse grouping the log hazard ratio is estimated only approximately, and in some reviews it has been referred to as a log odds ratio (Early Breast Cancer Trialists' Collaborative Group 1990). If the time intervals are large, a more appropriate approach is one based on interval-censored survival (Collett 1994).

The second approach can be used if trialists have analysed the data using a Cox proportional hazards model, or if a Cox model is fitted to individual patient data. Cox models produce direct estimates of the log hazard ratio and its standard error (so that a generic inverse variance meta-analysis can be performed). If the hazard ratio is quoted in a report together with a confidence interval or P value, estimates of standard error can be obtained as described in Section 7.7.7.